Assessment Information

CoreTrustSeal Requirements 2020–2023

Repository:	Zuse Institute Berlin - EWIG
Website:	ewig.zib.de
Certification period:	30 November 2023 - 29 November 2026
Requirements version:	CoreTrustSeal Requirements 2020-2022
This repository is owned by:	**Zuse Institute Berlin**

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background Information

Repository Type

Please provide context for your repository. You can select one or multiple options.

Compliance level:

Not Applicable - 0

Response:

- Archive
- Library
- Museum
- National repository system; including governmental
- Research project repository

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Description of Repository

Provide a short overview of the repository.

Compliance level:

Not Applicable - 0

Response:

The long-term archive has been given the name EWIG (formerly an acronym, which translates to "eternal" in english). EWIG is an OAIS-compliant long-term digital archive for cultural heritage and research data in Germany. EWIG is operated by the Zuse Institute Berlin (ZIB), a non-university research institute financed by the Federal State of Berlin.

Libraries, museums, archives and research institutions (depositors) have the possibility to archive data at ZIB for longer than the foreseeable future. Since the ZIB is a research institute, EWIG is designed to take care for research data as well. The retention time could be set to 10 years as required in the DFG Kodex (Guidelines of the German Research Foundation) or longer, depending on the relevance of the data. The ingest pipeline and requirements on preservation information data is not dependent on the agreed upon retention period in that regard that there is no low level way of archiving in EWIG.

EWIG acts as a dark archive and is not intended as a self service repository like zenodo or dryad. Each preservation project in EWIG starts with a consulting phase, where the intentions of the prospective customer are clarified. In that phase a briefing on available and recommended preservation metadata and formats takes place and typically there are a few iterations of the provided metadata.

Zuse Institute Berlin - EWIG

EWIG is run and developed by the Digital Preservation Working Group (AG DP) at ZIB. The working group is part of the Digital Data and Information for Society, Science, and Culture (D^2IS^2C) department.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Designated Community

Provide a clear definition of the Designated Community

Compliance level:

Not Applicable - 0

Response:

EWIG is open to libraries, archives and museums as well as scientific institutions to preserve their output for at least the recommended 10 years for scientific output, in case of cultural heritage even longer. The designated community consists at the present time solely in the institutional depositors themselves. Since EWIG represents a background service of ZIB for the depositors, there is no externally available presentation layer for end users (consumers). In the past there has been one archiving project where landing pages for the content have been provided by a co-operating working group parallel to the archiving of the assets but that was exclusively limited to that one project. EWIG is intended as a dark archive.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Level of Curation

Select all relevant types of curation.

- Content distributed as deposited
- Basic curation – e.g., brief checking, addition of basic metadata or documentation
- Enhanced curation – e.g., conversion to new formats, enhancement of documentation
- Data-level curation – as above, but with additional editing of deposited data for accuracy

Compliance level:

Not Applicable - 0

Response:

- A. Content distributed as deposited
- B. Basic curation – e.g. brief checking; addition of basic metadata or documentation
- C. Enhanced curation – e.g. conversion to new formats; enhancement of documentation

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Level of Curation - explanation

Please add the description for your Level(s) of Curation.

Compliance level:

Not Applicable - 0

Response:

A. Content distributed as deposited

Within normal operation not applicable except for longer phases of intermediate storage where depositors are able to prepare, amend and curate digital material until deposit for pre-ingest. The archive maintains (short-term) data integrity only. This level does not correspond to the ingest phase in which the archive assumes responsibility for the long-term.

B. Basic curation – e.g. brief checking; addition of basic metadata or documentation

Minimal checks of completeness, integrity and metadata description as required by policy. This includes at least a name or title, a creator/creation context, well-formedness and administrative information about deposit agreement, identification and rights as well as file identification, technical characterization, and validation.
This level applies to all data for which a transfer of custody is issued i.e. which are successfully ingested into the Long Term Preservation System.

C Enhanced curation – e.g. conversion to new formats; enhancement of documentation

Additional to basic level curation this level of curation includes more meticulous validation of structured metadata schema (LIDO, EAD, MODS) and mapping to Dublin Core Terms (DCT), a creator/creation context and administrative information about deposit agreement, deposit description, identification and rights. Restructuring based on content as master, metadata (original or normalized), auxiliary metadata, deposited intermediates (e.g. cropped or color-corrected images as derivatives), and different representations of material (e.g. automatic or manual transcriptions). This level usually depends on the contract with the depositor and involves a lot of communication, which can evolve into a project. Therefore this level is applied to less than 10 per cent of the data.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Insource/Outsource Partners

If applicable, please list them.

Compliance level:

Not Applicable - 0

Response:

The Digital Preservation Working Group (AG DP) at the ZIB is a joint effort of two D²IS²C research service units involved in preserving digital data in the long term: the Research and Competence Center Digitalization Berlin (digiS) which coordinates and supports digitialization efforts of cultural assets in the city state of Berlin, and the Cooperative Library Collaboration Berlin-Brandenburg (KOBV) which is the hub for all university libraries, all public libraries and numerous research, special and government libraries in the federal states of Berlin and Brandenburg.

https://www.digis-berlin.de/
https://www.kobv.de/

KOBV and digiS established in cooperation a research service group “Digital Preservation” that utilizes the available supercomputing infrastructure of the ZIB to implement a long-term digital preservation archival information system (DPS) that supports cultural heritage and research institutions. The DPS offers deposit, curation and preservation services for any kind of data that data producers would want to keep safe-guarded.

https://www.zib.de/disc/digital-preservation

https://www.zib.de/research_services/supercomputing

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Significant Changes

Summary of Significant Changes Since Last Application if applicable.

Compliance level:

Not Applicable - 0

Response:

-

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Other Relevant Information

You may provide other relevant information that is not covered by the requirements.

Compliance level:

Not Applicable - 0

Response:

EWIG is designed as a dark archive. Access is limited to the digital archivists for administrative purposes and to the depositor. EWIG has a regional and, to a lesser extent, national focus.

Currently EWIG holds approximately 100 collections, archived from 30 Institutions. Roughly one third of this holdings are for 10 years, the rest has an indefinite retention time.

The archive serves as an Long Term Preservation (LTP) solution for the depositors. The presentation of the access copies for end-users is provided via their live systems (e.g. Digital Asset Management (DAM), repositories, data catalogs, etc.). The responsibility for access remains with the depositors. EWIG secures the long term availability and (usability) of the digitized raw sources and accompanying metadata. In case the depositor has obsolete access copies, EWIG provides him with state of the art access derivatives.

Links:

Reviews

Reviewer 1:

Compliance level:

Not Applicable - 0

Comments:

Reviewer 2:

Compliance level:

Not Applicable - 0

Comments:

Organizational Infrastructure

R1 Mission/Scope

The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

The two comprising bodies of EWIG (KOBV and digiS) have both an explicit mandate from the state of Berlin to provide Long Term Preservation services for the states of Berlin and Brandenburg.

For the KOBV it is expressed in the service specifications (Leistungsverzeichnis in German) which can be found under the following link: https://www.kobv.de/kobv/verbund/beitragsordnung/

digiS has been established in 2014 as a central coordinating office for digitizing and preserving the cultural heritage of Berlin. The mandate can be found under this link: https://www.berlin.de/sen/kultur/en/cultural-policy/cultural-participation/digitalization/digitization-of-assets/

The ZIB department Digital Data in Science Society and Culture (D^2IS^2C) which is represented by the aforementioned bodies KOBV an digiS has its main goal in planning and developing of information systems and services for scholarly research and cultural heritage institutions. Services are provided for

- Mathematical institutions (International Mathematical Union IMU and Math-Net)
- Libraries (KOBV)
- Museums (Rearch Group Information Technology Tools for Museums)
- Archives
- and for a variety of Digitization and Research Data Management projects and Digital Preservation efforts (digiS, Research Group Digital Preservation) http://www.zib.de/disc

More information about the service of digital preservation is published online. It can be seen as the central mission of the ZIB with regard to digital archiving:

“Our main task is to develop and operate a digital preservation system (long-term archive) to provide permanent access to the results of scholarly research and to the outcome of digitization efforts from a wide variety of cultural heritage institutions. […] The idea is to distinguish the domain of active research and the long-term preservation domain. In the active research domain, curation is the responsibility of the researcher or heritage institution, and in the long-term preservation domain, responsibility lies with the long-term archive. An important part of the work is focused on data curation issues, i.e., a thorough look inside the transferred data stream to secure file integrity and authenticity.”
http://www.zib.de/disc/digital-preservation

A thorough explanation of services and approaches of EWIG can be found on the homepage of the archive itself: https://ewig.zib.de/service/
(All the information there is momentarily in German language only. EWIG is primarily provided as a service to the states of Berlin and Brandenburg. Albeit it is possible to come to arrangements with customers from other countries, there is no focus on providing international services)

Information about the central digital preservation service is promulgated via relevant websites and networks:

Kooperativer Bibliotheksverbund Berlin-Brandenburg:
https://www.kobv.de/services/archivierung/lza/

Guidance Policy of the Research and Competence Center Digitalization Berlin (digiS) :
https://dx.doi.org/10.12752/2.0.003.1

The ZIB itself has a policy for handling research data which takes the long term availability of research data into account:
https://www.zib.de/sites/default/files/guidelines/ZIB-Research-Data-Policy-20190306.pdf

Links:

- https://www.kobv.de/kobv/verbund/beitragsordnung/
- https://www.berlin.de/sen/kultur/en/cultural-policy/cultural-participation/digitalization/digitization-of-assets/
- http://www.zib.de/disc
- http://www.zib.de/disc/digital-preservation

- https://ewig.zib.de/service/
- https://www.zib.de/sites/default/files/guidelines/ZIB-Research-Data-Policy-20190306.pdf
- https://www.kobv.de/services/archivierung/lza/
- https://dx.doi.org/10.12752/2.0.003.1

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R2 Licenses

The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

The Zuse Institute Berlin (ZIB) employs persons specialized in legal issues, privacy, data security, and negotiations of contracts that support and monitor the work of the institute's staff, and also that of the research service group "Digital Preservation".
https://www.zib.de/institute/administration

Prior to all service related interactions all institutions who wish to deposit with us have to sign a legal contract and an accompanying contract annex regulating compliance with the General Data Protection Regulation (EU). This contract lays down the terms of access and the rights management. A model contract is accessible. The contract states e.g. the rights of the depositor to hand over the data to EWIG and on the side of EWIG prohibits the Archive from handling the data outside of the granted permissions. If the conditions of this contract are not met on the side of the depositor, there will be no service provided by EWIG. If EWIG violates the contract while there is a service relationship, it may be held liable for the damage incurred.

Our DPS is designed as a dark archive. Access is limited to the digital archivists for administrative purposes and to the depositor. Producers have to supply a mandatory submission manifest that provides details about access rights such as copyright, data protection, personal rights, embargo periods, authenticated persons, and license information.

Our archive is also governed by the legal statements of its host institution, the ZIB, a public institution established under its own law with the task to operate in close cooperation with the universities and scientific institutions in Berlin for the advancement of research and development in the field of information technology and to meet the related service needs (see § 2 of the on-line available legal text):
https://www.zib.de/institute/organization/law

Links:

- https://www.zib.de/institute/administration
- https://www.zib.de/institute/organization/law

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R3 Continuity of access

The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance level:

The repository is in the implementation phase - 3

Response:

Since the DPS serves as a third party archive, a contract between the customer (or depositor) and the archive is signed. Every depositor pays for the desired time he wants to store the data. This period is usually 5 years, 10 years, or indefinite. For this time the archive takes full responsibility of maintaining its operation, since all operational costs are covered.

In case that the DPS is no longer capable of maintaining its core duties, the archived assets can be returned to the customer or transferred to an alternative archive. For that purpose an uncompressed TAR archive of every Archival Information Package is built and will be staged for the customer to download.

At this point of time there is no written agreement between ZIB and the financing Federal State of Berlin regarding transfer of responsibility for the assets. Therefore we apply for compliance level 3.

Links:

Reviews

Reviewer 1:

Compliance level:

The repository is in the implementation phase - 3

Comments:

Reviewer 2:

Compliance level:

The repository is in the implementation phase - 3

Comments:

R4 Confidentiality/Ethics

The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

EWIG is not a discipline-specific repository serving third-party consumers and has no control over how research- or cultural heritage data is generated. The data is returned exclusively to the institution that has the archiving contract with EWIG. Therefore the responsibility for discipline-specific standards of access and re-use of the archival records remains with the respective repositories of the data-providing institution.

Before submission of the data a deposit agreement is negotiated between the producer and EWIG. It comprises a legal contract for transfer of custody including responsibilities of producer and the Digital Preservation System, as well as a technical policy (submission agreement) giving more specific advice and guidance for data producers on metadata and documentation that should be associated with the data. The submission agreement is based on the standard DIN 31645 “Guide to the transfer of information objects into digital long-term archives”.

Link to submission guidelines of EWIG: https://ewig.zib.de/static/submission_guidelines.html

Links:

- https://ewig.zib.de/static/submission_guidelines.html

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R5 Organizational infrastructure

The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

EWIG is located at the Zuse Institute Berlin (ZIB). The ZIB is a research institute for applied mathematics and computer science and operates a regional supercomputing center that requires in-house facilities to store more than five petabytes of data on disk and nearly a hundred petabytes on tape.
https://www.zib.de/research_services/supercomputing

The Digital Preservation Working Group (AG DP) is a joint effort of two ZIB research service units involved in preserving digital data in the long term: The Research and Competence Center Digitalization Berlin (digiS) which coordinates and supports digitialization efforts of cultural assets in the city state of Berlin, and the Kooperativer Bibliotheksverbund Berlin-Brandenburg (KOBV) which is the network of all university libraries, all public libraries and numerous research, special and government libraries in the federal states of Berlin and Brandenburg.

https://www.digis-berlin.de/
https://www.kobv.de/

KOBV and digiS established in cooperation a research service group “Digital Preservation” that utilizes the available supercomputing infrastructure of the ZIB to implement a long-term digital preservation archival information system (DPS) that supports cultural heritage and research institutions. The DPS offers deposit, curation and preservation services for any kind of data that data producers would want to keep safe-guarded.

https://www.zib.de/disc/digital-preservation

The DPS is trustworthy in the sense that it is transparent in its functionality, documentation, and policy and is aligned to the Open Archival Information System (OAIS) reference model.

EWIG brings 3 FTE positions. The competences are distributed between software developers, metadata specialists, a quality assurance manager and a supervising manager. The tasks are not always strictly separable, so a software developer could act as metadata analyst ore vice versa and quality assurance could act as developer. The team of EWIG attends the relevant international and national conferences. Three of the team-members are active in nestor working groups. In addition, the EWIG team members of KOBV and digis have a regular budget for both ongoing training and professional development in IT and information management.
All Team members and all supporting IT staff managing the IT-Infrastructure are on permanent contracts. All funding of EWIG is structural, not project based.

Links:

- https://www.zib.de/research_services/supercomputing
- https://www.digis-berlin.de/
- https://www.kobv.de/
- https://www.zib.de/disc/digital-preservation

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R6 Expert guidance

The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either in-house, or external, including scientific guidance, if relevant).

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

There is a Scientific Advisory Board that evaluates all areas of ZIB's work once a year and the Long Term Preservation System is part of that. There also is a specialised Advisory Board with members of the library, museum, archival and information management community which annually meet at ZIB and provides feedback regarding cultural Heritage collections and projects in custody of EWIG.

EWIG-Team Members participate regularly in the iPRES and RDA-Conventions.

digiS is heavily involved in building LTP competence in the respective designated community. To this end it holds workshops and cooperates with various institutions and organizations on the different aspects of reusability and availability of digital cultural heritage data. More information on workshops in German can be found here: https://www.digis-berlin.de/digis/veranstaltungen/workshops/

digiS advises cultural institutions, coordinates the digitization funding program, and provides technical services and expertise to ensure the long-term reusability and availability of the data from the projects. Here you can find, again in German only, ressources provided by digiS: https://www.digis-berlin.de/wissenswertes/

Links:

- https://www.digis-berlin.de/wissenswertes/
- https://www.digis-berlin.de/digis/veranstaltungen/workshops/

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Digital Object Management

R7 Data integrity and authenticity

The repository guarantees the integrity and authenticity of the data.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

Under https://ewig.zib.de/static/datamodel.html information is provided on the EWIG datamodel. Since this is technical Information from and for developers it is readily available in English.
The description of all entities involved in the preservation process and their relations are described in detail.

EWIG goes through some lengths to provide customers / depositors prior to actually putting their assets in EWIG with a thorough understanding of the terms and processes involved in Long Term Preservation. This stems in part from the capacity building mission of digiS. If there is no standard transfer package already established at the depositor, EWIG aims at Metadata and structural description in the Metadata Encoding & Transmission Standard (METS) maintained by the Library of Congress under https://www.loc.gov/standards/mets/

To this end EWIG provides its customers with a detailed description (alas again in German only) of the expected METS representation under: https://ewig.zib.de/static/METS-DRAFT.html#alternative-lieferung-mehrerer-ies-als-einzelne-mets-dokumente

EWIG uses the Archivematica engine for its core functionality.
Citing their Website under: https://www.archivematica.org/en/
"Archivematica is an integrated suite of open-source software tools that allows users to process digital objects from ingest to access in compliance with the ISO-OAIS functional model. Users [can] monitor and control ingest and preservation micro-services via a web-based dashboard [or deploy the service via API calls.]. Archivematica uses METS, PREMIS, Dublin Core, the Library of Congress BagIt specification and other recognized standards to generate trustworthy, authentic, reliable and system-independent Archival Information Packages (AIPs) for storage in your preferred repository."

The Ingest pipeline is described (in German) under: https://ewig.zib.de/archivierungsworkflow/#Ingest

All actions undertaken in the archive are documented in the PREMIS (Data Dictionary for Preservation Metadata) section of the METS accompanying the Information Package. PREMIS is another Standard maintained by the Library of Congress and Information on how to use it in conjunction with the aforementioned METS can be found under: https://www.loc.gov/standards/premis/premis-mets.html

Links:

- https://ewig.zib.de/static/datamodel.html
- https://www.loc.gov/standards/mets/
- https://ewig.zib.de/static/METS-DRAFT.html#alternative-lieferung-mehrerer-ies-als-einzelne-mets-dokumente
- https://www.archivematica.org/en/
- https://ewig.zib.de/archivierungsworkflow/#Ingest
- https://www.loc.gov/standards/premis/premis-mets.html

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R8 Appraisal

The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

In a subsection to the complete description of the archiving workflow (available only in German) under: https://ewig.zib.de/archivierungsworkflow/ information can be found in two places on the Pre-Ingest and Ingest itself.

All necessary prerequisites from the data depositor are listed under: https://ewig.zib.de/archivierungsworkflow/#Pre_Ingest and under https://ewig.zib.de/archivierungsworkflow/#Ingest with a list of preferred data formats and the likeliness of satisfactory preservation action that could be performed on them.

Deposited information is preserved as self contained and self describing archival information packages (AIP) using the preservation metadata dictionary as described by PREMIS (https://www.loc.gov/standards/premis/).

The data producer is obliged to supply administrative metadata (= preservation description information, PDI) that contains information about the producer himself as well as identifiers, contract numbers, contacts, rights etc. The DPS staff gives support and controls the quality of the PDI. A template of the submission manifest is given in the submission guideline in German: https://ewig.zib.de/static/submission_guidelines.html

The provision of descriptive metadata (descriptive information, DI) is mandatory. Supported standardized metadata formats are Encoded Archival Description (EAD) for deposits from archives, Lightweight Information Describing Objects (LIDO) for museums, Metadata Object Description Schema (MODS) for libraries and qualified DC-XML. These formats are automatically extracted and mapped to Dublin Core (DC) by the EWIG ingest pipeline.

https://www.loc.gov/ead/
https://cidoc.mini.icom.museum/working-groups/lido/lido-overview/about-lido/what-is-lido/
https://www.loc.gov/standards/mods/

Further documentation can be added which might be useful for understanding the deposited material in the future but is not part of the information object. During the ingest process technical metadata is extracted and all actions taken are documented.

The tools FIDO, Siegfried and File Extension are used for format identification. FIDO & Siegfried identify files via embedded signatures. File Extension only examines the file format extension. File Extension can also be used for formats that do not have a PRONOM ID (PUID).
In the format identification, technical metadata is extracted - for image files, for example, the exif information. The tools exiftool, MediaInfo, FITS and FFIdent are used. For file format validation Archivematica relies on the standard tool JHOVE. Additionaly we use checkitTIFF, since it has some advantages over JHOVE.

All metadata submitted by the producer and automatically generated is organized into a METS file (Metadata Encoding and Transmission Standard - https://www.loc.gov/standards/mets/). Together with the information object, it is compiled into a bag-it structure (https://datatracker.ietf.org/doc/html/rfc8493) that is saved as a single archive file to allow for easy transfer within the data management layer.

There is a defined procedure for the removal of items (Löschauftrag in German) from the collections. Since EWIG is a dark archive this request for removal must come from the depositor.

Links:

- https://www.loc.gov/ead/
- https://datatracker.ietf.org/doc/html/rfc8493
- https://ewig.zib.de/archivierungsworkflow/
- https://ewig.zib.de/archivierungsworkflow/#Pre_Ingest
- https://www.loc.gov/standards/premis/
- https://www.loc.gov/standards/mets/
- https://ewig.zib.de/static/submission_guidelines.html
- https://ewig.zib.de/archivierungsworkflow/#Ingest

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R9 Documented storage procedures

The repository applies documented processes and procedures in managing archival storage of the data.

Compliance level:

The repository is in the implementation phase - 3

Response:

All the technical Infrastructure of EWIG is located in the ZIB building itself, so there is no need for specialized agreements with external institutions, since all the processes are documented in the IT Organizational and Operational Management (ITOOM) Document from the IT department. All data goes automatically on two tapes in two distinct tape libraries and processes are in place to verify the fixity information of the stored files to ensure data integrity. Also there is a policy in place for the tape library which checks for the amount of access events and the overall age of the tapes. Whenever one of the defined thresholds is reached (age or number of access events) the storage components (tapes) are replaced.

A ZIB internal department called “Datacenter Infrastructure” is responsible for continuous operation and maintenance of the storage infrastructure (https://www.zib.de/itds).

The tape libraries are installed in a special vault that is water proof, can withstand an outside fire for around 10 hours, and has an additional CO2-fire extinguisher system.

There is a very detailed complete data flow diagram produced by the working group digital preservation. Since it contains sensitive infrastructure information according to the colleagues from the IT department it is not freely available at the moment. We are working on a publishable version for our next application.

EWIG does a lot of the ingest development inhouse. There is a GitLab instance installed on premise, where development takes place. A lot of documentation is provided to EWIG staff via the corresponding GitWiki. This documentation covers the necessary steps for transferring data packages from the customer upload area through ingest and processing via Archivematica up to final storage on the tape library. Topics covered include

- Copy from sftp upload area
- Copy from tape
- Check integrity
- Restructure data according to EWIG guidelines if necessary
- Ingest on test-pipeline
- Ingest on production-pipeline
- Possible errors during Archivematica-processing

For each of the topics commands and procedures are stated and documented so that even administrative staff will be able to handle the archiving process. Since this documentation is available only to EWIG staff from within the ZIB no precautions have been applied to mask sensitive information. There are machine names, paths and other critical infrastructure information, which should not be available outside the ZIB. Like with the workflow diagram we intend to have a more abstract version available for publication with our next application.

Links:

- https://www.zib.de/itds

Reviews

Reviewer 1:

Compliance level:

The repository is in the implementation phase - 3

Comments:

Reviewer 2:

Compliance level:

The repository is in the implementation phase - 3

Comments:

R10 Preservation plan

The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance level:

The repository is in the implementation phase - 3

Response:

On https://ewig.zib.de/service/ there is a diagram of the preservation workflow in English language publicly available at "Architektur des Systems" (Architecture of the System) . Information on preferred formats and actions which can be performed on them is available under: https://ewig.zib.de/archivierungsworkflow/#Ingest

Our DPS supports the preservation of any data regardless of format. For identifiable data formats the preservation system assures renderability through migration. Data that are deposited in non-identifiable formats will be preserved on the bit-level to retain structure and metadata. The choice of preservation format is based on community best practices, availability of open-source normalization tools, and an analysis of the significant characteristics for each media type.

Preservation actions that migrate file formats are in place. Ingest components of our core preservation system Archivematica add features that allow for automatic re-ingestion and normalization of AIPs. The supplied or extracted technical metadata is examined through preservation watch. Format migration to stop future obsolescence will only be carried out if there are relevant changes in the format policy. These migrations are documented within the archives management system creating a record of all processes enacted on files, allowing archivists to manage current and future migrations.

Links:

- https://ewig.zib.de/archivierungsworkflow/#Ingest
- https://ewig.zib.de/service/

Reviews

Reviewer 1:

Compliance level:

The repository is in the implementation phase - 3

Comments:

Reviewer 2:

Compliance level:

The repository is in the implementation phase - 3

Comments:

R11 Data quality

The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality- related evaluations.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

In order to preserve as much contextual information as possible and to make the content understandable, depositors are required to provide mandatory descriptive metadata for the data objects according to the specification under: https://ewig.zib.de/static/METS-DRAFT.html. Supplementary technical, administrative, and preservation metadata will be added by EWIG as PREMIS metadata during the ingest process.

In a submission manifest, depositors must include human- and machine-readable information on e.g. provenance, responsible contact persons, and third-party access rights for the deposited data.The submission guidelines with an examplary submission manifest (in German only) can be found under: https://ewig.zib.de/static/submission_guidelines.html

The adherence to standards of data formats is ensured by standard processes implemented in Archivematica for characterization, identification and validation where feasible. EWIG requires the delivery of curated data, as no normalization-on-ingest is performed. In case of problems detected while ingesting deposited data, the transfer curator is contacted for options such as rectification and redeposit or to authorize the ingest despite the shortcomings.

Depositors are encouraged to provide additional information describing the deposits via an auxiliary "submission documentation". This can be e.g. insurance data of museum objects, e-mails or photos of the setup of a research experiment.

Links:

- https://ewig.zib.de/static/METS-DRAFT.html
- https://ewig.zib.de/static/submission_guidelines.html

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R12 Workflows

Archiving takes place according to defined workflows from ingest to dissemination.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

Our OAIS aligned preservation system combines existing free and open-source components including Archivematica (digital preservation system), iRODS (data management software), and Fedora (repository system). An additional triple store, which is together with Fedora accessed by a custom API, is used for administrative data management. Integrating different components into one system offers the opportunity for clear responsibilities, audit, and documentation and therefore trustworthiness.

The DPS as a whole consists of the stages “preingest”, “ingest”, “management”, and “access”. Only the first and the last stage, preingest and access, are customized to meet the requirements of different types of data, the core stages (ingest and management) process the data from all producers the same way.
https://www.archivematica.org/en/
https://irods.org/
https://fedora.lyrasis.org/

The whole system design and implementation is regarded as a living system that has to be cared for and adapted to a changing environment. We try to achieve this goal by utilizing a single core preservation workflow for all information packages to reduce complexity and increase sustainability.
Our preservation system accepts any identifiable data formats of digital material for deposit. For these data the preservation system assures the functioning of the described archiving workflow and renderability through migration. Data that are deposited in non-identifiable formats will be preserved on the bit-level.

A detailed textual description of the preservation workflow (in german) is available for data producers: https://ewig.zib.de/archivierungsworkflow/ An english diagram on alignment of the various components of EWIG is available at: https://ewig.zib.de/service/

Links:

- https://ewig.zib.de/archivierungsworkflow/
- https://irods.org/
- https://fedora.lyrasis.org/
- https://www.archivematica.org/en/
- https://ewig.zib.de/service/

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R13 Data discovery and identification

The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance level:

The repository is in the implementation phase - 3

Response:

After transmission of the Archival Information Packages (AIP) to the archival storage the Digital Preservation System (DPS) takes over responsibility for the preservation and long term accessibility of the digital content. At this stage and for documentation purposes only, a submission report is prepared and submitted to the depositor.

All stored digital assets are addressable to the team of EWIG via the so called reporting dashboard (developed inhouse) based on the Fuseki Triple Store. Each item is assigned a Universally Unique Identifier (UUID) in the ingest process, which stays permanent in the EWIG context. Stored items can be searched and identified through various filters like depositor, time of ingest, name of the submission, name of the Intellectual Entity itself, etc. See https://ewig.zib.de/archivierungsworkflow/#Data_Management (German only). At the moment the reporting dashboard is available only to the EWIG Team but development is underway to provide read only access for the depositors to the stored assets for statistical overview information. EWIG does not plan to become a patron managed LTP System in the foreseable future.

The transfer of custody including responsibilities of producer and the DPS forms part of the legal agreements between digital archivists and institutional depositor. The archive ensures access and availability of the digital objects for the depositor. Access to Dissemination Information Packages (DIP) is only provided for the depositor (dark archive). Responsibility for (online) presentation of digital collections remains with the depositor.

Links:

- https://ewig.zib.de/archivierungsworkflow/#Data_Management

Reviews

Reviewer 1:

Compliance level:

The repository is in the implementation phase - 3

Comments:

Reviewer 2:

Compliance level:

The repository is in the implementation phase - 3

Comments:

I agree with Level of compliance 3 as suggested by the applicants.

R14 Data reuse

The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

EWIG emphasizes a rich metadata set, regardless of the "user" of the data. As a dark archive we do not provide access for third party users. But even if it is the data depositor itself, a rich metadata set is of high value. Be it that the repository software at the depositors premises has changed or a new way of cataloguing has emerged.
The publicly available submission guidelines (in German) define the required administrative metadata including rights, licenses, access rights and additional metadata : https://ewig.zib.de/static/submission_guidelines.html#notwendige-metadaten-der-datenlieferung

The repository is currently only accessible for internal administrative users. Access to the AIPs for data producers is realized with management actions by the administrative staff to stage data on a location where it can be picked up; we provide no self-service at the moment.
https://ewig.zib.de/archivierungsworkflow/#Access

With changing requirements of our clients/depositors we might consider to implement an access repository for bulk self-service AIP exports and DIP-download in the future.

In the past there has been a project for a client that had no means of providing access to their data themselves. We offered a service for generating customized, static landing pages for each DIP as a low maintenance way to presenting them on the web.

Links:

- https://ewig.zib.de/static/submission_guidelines.html#notwendige-metadaten-der-datenlieferung
- https://ewig.zib.de/archivierungsworkflow/#Access

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Technology

R15 Technical infrastructure

The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

Our DPS uses an OAIS aligned data model and architectural design that enables us to archive digital information with a single core preservation workflow. The system combines existing free and open-source components including Archivematica, iRODS, and Fedora. Because of our requirements to be aligned to OAIS and have self-contained AIPs for interoperability, we designed the architecture to be a single, modular core pipeline of existing tools linked up by in-house developed data conduits. The system is therefore composed of loosely coupled components with strictly defined responsibilities. As we anticipate the software components to become obsolete during the lifetime of the system, this modularity enables us to find or develop tools to substitute waning functionalities. The whole system design and implementation is regarded as a living system that has to be cared for and adapted to a changing environment and new requirements. We try to achieve this goal by utilizing a single core preservation workflow for all information packages to reduce complexity and increase sustainability.
Our in-house developed software components are managed in git repositories and mainly deployed with Ansible on test and production systems.
The architectural entities run on dedicated Virtual Machines with current versions of Ubuntu, CentOS and RedHat operating systems which are regularly backed up.

Persistent storage consists of a hierarchical Storage Archive Manager that augments an online file system transparently with nearline redundant tape storage. ZFS is used for the online filesystem that is designed to prevent data corruption caused by bit-rot. The data is stored redundantly on two StorageTek 8500 tape libraries with currently installed tape capacity of around 100 petabyte, of which nearly 400 terabyte (800 terabyte with redundancy) are currently reserved for our archival system. The libraries have no physical connection to prevent tapes from being destroyed by tape recorder malfunctions. They also use automatic fixity checks and error correction to ensure data integrity.

See R16 for further information regarding the technical infrastructure.

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

R16 Security

The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

EWIG is hosted on premise on the ZIB own IT infrastructure. Access to the virtual machines comprising the LTP System is granted to IT-Admins of ZIB and members of the EWIG team only from within the ZIB-network via password or key-based authentication or in case of the reporting dashboard via username and password. The only part accessible from outside the ZIB-network at the moment is the depositor's individual staging area or "transfer folder" via sftp and key-based authentication which is served by a separate virtual machine.

The IT department has developed an IT security guideline which covers all areas of data security. Citing from the guideline:
Based on the recommendations and suggestions of the German Federal Office for Information Security (BSI) and the provisions of the General Data Protection Regulation (GDPR), and in order to meet the legal and contractual requirements as well as the targets we have set ourselves, it is necessary to,

- implement suitable and appropriate measures to prevent incidents and to maintain the functionality of the IT infrastructure,
- recognize that regular action is required to continue to ensure the long-term and sustainable security of research data, data from cooperation partners, and compliance with data protection regulations,
- safeguard the rights of users to their personal data.

ZIB's IT security policy is designed to meet these requirements and is based on the principles set forth by the BSI in the IT-Grundschutz, as well as the DSGVO, the German Federal Data Protection Act (BDSG), and the Berlin Data Protection Act (BlnDSG), including "Best Practices in Research and Higher Education".

The IT security policy of the Zuse Institute Berlin describes

- the handling of IT procedures with regard to information and data security,
- the responsibility of managers and employees in the deployment and use of IT procedures,
- the regulations for administrators and users of IT procedures to comply with information and data security,
- procedures for classifying the need for protection of IT processes and the resulting protective measures

Security and data protection in information technology serve to ensure the availability, integrity and confidentiality of data and IT applications. When processing personal data, the principles of data avoidance, data economy and necessity, and purpose limitation must also be observed. They can only be achieved through a bundle of measures from the areas of organization, personnel, infrastructure, hardware and software, communication and emergency preparedness.

List of covered topics:

IT organization (M01 - M30)
IT personnel (M31 - M60)
Securing the infrastructure (M61 - M90)
Hardware and software deployment (M91 - M120)
Deployment of mobile devices (M121 - M150)
System and network management (M151 - M180)
Access protection (M181 - M210)
Data backup (M211 - M240)
Data medium control (M241 - M270)
Logging (M271 - M300)
Basic protection for IT users (M301 - M330)

The tape libraries are installed in a special vault that is water proof, can withstand an outside fire for around 10 hours, and has an additional CO2-fire extinguisher system.

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Applicant Feedback

R17 Applicant Feedback

We welcome feedback on the CoreTrustSeal Requirements and the Certification procedure.

Compliance level:

The guideline has been fully implemented in the repository - 4

Response:

Dear Reviewers, dear Board,
thanks again for taking the time for thorough reading. Regarding the suggestion of one reviewer to raise the desired level of compliance to 4 at requirement 10 we would like to stick to 3 at the moment and enrich the documentation for our next CTS application, maybe aiming for a 4 then.
Best from Berlin,
Tim

Links:

Reviews

Reviewer 1:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments:

Reviewer 2:

Compliance level:

The guideline has been fully implemented in the repository - 4

Comments: